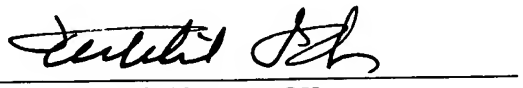
I, Tadahiko Itoh, a Patent Attorney of Tokyo, Japan having my office at 32nd Floor, Yebisu Garden Place Tower, 20-3 Ebisu 4-Chome, Shibuya-Ku, Tokyo 150-6032, Japan do solemnly and sincerely declare that I am the translator of the attached English language translation and certify that the attached English language translation is a correct, true and faithful translation of Japanese Patent Application No. 11-118072 to the best of my knowledge and belief.

I hereby declare that all statements made herein of my own knowledge are true and that all statements made on information and belief are believed to be true; and further that these statements were made with the knowledge that willful false statements and the like so made are punishable by fine or imprisonment, or both, under Section 1001 of Title 18 of the United States Code and that such willful false statements may jeopardize the validity of the application or any patent issued thereon.



Tadahiko ITOH
Patent Attorney
ITOH International Patent Office
32nd Floor,
Yebisu Garden Place Tower,
20-3 Ebisu 4-Chome, Shibuya-Ku,
Tokyo 150-6032, Japan

(Document Name)	Application for Patent
(Reference Number)	NTTH107475
(Date of Submission)	April 26, 1999
(Destination)	Commissioner of Patent Office
(IPC)	G06C
(Inventor)	
(Residence or Address)	C/O NIPPON TELEGRAPH AND TELEPHONE CORPORATION 19-2 Nishi-Shinjuku 3-chome Shinjuku-ku, Tokyo, Japan
(Name)	Hiroshi Ogawa
(Inventor)	
(Residence or Address)	C/O NIPPON TELEGRAPH AND TELEPHONE CORPORATION 19-2 Nishi-Shinjuku 3-chome Shinjuku-ku, Tokyo, Japan
(Name)	Takao Nakamura
(Inventor)	
(Residence or Address)	C/O NIPPON TELEGRAPH AND TELEPHONE CORPORATION 19-2 Nishi-Shinjuku 3-chome Shinjuku-ku, Tokyo, Japan
(Name)	Atsuki Tomioka
(Inventor)	
(Residence or Address)	C/O NIPPON TELEGRAPH AND TELEPHONE CORPORATION 19-2 Nishi-Shinjuku 3-chome Shinjuku-ku, Tokyo, Japan
(Name)	Youichi Takashima
(Applicant for Patent)	
(Identification Number)	000004226
(Name)	NIPPON TELEGRAPH AND TELEPHONE CORPORATION
(Attorney)	
(Identification Number)	100066153
(Patent Attorney)	
(Name)	Takashi Kusano

(Designated Attorney)
(Identification Number) 100100642
(Patent Attorney)
(Name) Minoru Inagaki
(Priority Claim based on Previous Application)
(Filing Number) Japanese Patent Application No.10-251193
(Filing Date) September 4, 1998
(Indication of Official Fees)
(Prepayment Ledger Number) 002897
(Amount Paid) ¥21,000
(Lists of Submitted Documents)
(Document Name) Specification 1
(Document Name) Drawing 1
(Document Name) Abstract 1
(Number of General Power of Attorney) 9806848
(Proof Requested or Not) Requested

[Name of the Document] Specification

[Title of the Invention]

METHOD AND APPARATUS FOR EXTRACTED DIGITAL WATERMARK STATISTICAL PROCESSING

[Claims]

1. A method for embedding digital watermark data in digital data contents, said method comprising the step of:

reconstituting said digital watermark data embedded in digital data contents from said bit sequence extracted from digital data contents before being reconstituted by using a test method on the basis of binary distribution in statistics.

2. The method as claimed in claim 1, further comprising the steps of:

presetting a reliability threshold value α of the extracted digital watermark data;

obtaining a binary distribution function F(x) which represents a probability that a number x of '1' bits or '0' bits are included in a bit sequence which is read at random from digital data contents, said binary distribution function F(x) being obtained by using a probability q of reading '1' or '0' in said bit sequence and a repeating number of embedding each bit of digital

watermark data;

reading an ith digital watermark sequence of said digital watermark data from a digital watermark area of said digital data contents;

calculating the number k_i of '1' or '0' included in said digital watermark sequence;

calculating a probability $F(k_i)$ by using said binary distribution function $F(x)$; and

reconstituting '1' or '0' from ith digital watermark data w_i if $F(k_i) > \alpha$, reconstituting '0' or '1' from ith digital watermark data w_i if $1-F(k_i) > \alpha$, and determining that there is no watermark data or the presence is unknown if both of $F(k_i) > \alpha$ and $1-F(k_i) > \alpha$ are not satisfied.

3. The method as claimed in claim 2, further comprising the steps of:

outputting $F(k_i)$ as reliability if said reconstituted digital watermark data w_i is '1'; and

outputting $1-F(k_i)$ as the reliability if said reconstituted digital watermark data w_i is '0'.

4. The method as claimed in claim 1, further comprising the steps of:

setting a reliability threshold value α of the

extracted digital watermark data;

obtaining a binary distribution function F(x) which represents a probability that a number x of '1' bits or '0' bits are included in a bit sequence which is read at random from digital data contents, said binary distribution function F(x) being obtained by using a probability q of reading '1' or '0' in said bit sequence and a repeating number of embedding each bit of digital watermark data;

reconstituting digital watermark data from said digital watermark sequence by using majority decision processing if said probability exceeds α, and determining that there is no watermark data or the presence is unknown if said probability does not exceed α.

5. The method as claimed in claim 4, further comprising the step of:

outputting said probability that said digital watermark sequence is digital watermark data as reliability of said reconstituted digital watermark data.

6. The method as claimed in any of the claims 1-5, if a data sequence which is embedded as said digital watermark data is modulated by a pseudo-random

sequence, said method further comprising the steps of:

demodulating said modulated data sequence by said pseudo-random sequence before the digital watermark data is reconstructed.

7. The method as claimed in claim 1, if a data sequence which is embedded as said digital watermark data is modulated by a pseudo-random sequence, said method further comprising the steps of:

setting a reliability threshold value α of the extracted digital watermark data;

obtaining a binary distribution function F(x) which represents a probability that a number x of '1' bits or '0' bits are included in a bit sequence which is read at random from digital data contents, said binary distribution function F(x) being obtained by using a probability q of reading '1' or '0' in said bit sequence and a repeating number t of embedding each bit of digital watermark data;

demodulating said modulated data sequence by said pseudo-random sequence before the digital watermark data is reconstructed;

assigning 1/2 to said probability q;

obtaining a maximum number x_0 which satisfies $0 \leqq F(x=x_0) \leqq 1-\alpha$ and a minimum number x_1 which satisfies $\alpha \leqq$

$F(x=x_1) \leqq 1, \alpha$;

obtaining the number k_i of '1' or '0' included in said ith digital watermark sequence; and

reconstituting ith digital watermark data w_i as '0' or '1' if $k_i \leqq x_0$, and, reconstituting said ith digital watermark data w_i as '1' or '0' if $k_i \geqq x_1$.

8. The method as claimed in claim 1, if a data sequence which is embedded as said digital watermark data is modulated by a pseudo-random sequence, said method further comprising the steps of:

demodulating said modulated data sequence by said pseudo-random sequence before the digital watermark data is reconstructed;

setting a reliability threshold value α of the extracted digital watermark data;

obtaining a binary distribution function F(x) which represents a probability that a number x of '1' bits or '0' bits are included in a bit sequence which is read at random from digital data contents, said binary distribution function F(x) being obtained by using a probability q of reading '1' or '0' in said bit sequence and a repeating number t of embedding each bit of digital watermark data;

assigning 1/2 to said probability q;

obtaining x_0 or x_1 which satisfies $0 \leqq F(x=x_0) \leqq 1-\alpha$ or $\alpha \leqq F(x=x_1) \leqq 1$;

judging whether the average value of the bias from the center of the binary distribution of the watermark sequence is equal to or more than x_1;

reconstituting digital watermark data by performing majority decision processing for said watermark sequence if said value is equal to or less than x_0 or equal to or more than x_1; and

determining that there is no digital watermark or the presence is unknown if said value is not equal to or less than x_0 or equal to or more than x_1.

9. The method as claimed in claim 8, further comprising the step of outputting said average value of the bias from the center of the binary distribution of the watermark sequence as reliability of said reconstituted digital watermark data in case the digital watermark data is reconstructed.

10. A method for reconstituting digital watermark data embedded in digital data contents, said method comprising the step of:

obtaining digital watermark data from the bias of appearance probability of the digital watermark data in

the binary distribution of appearance probability of each bit of 1 bit sequence extracted at random from digital data contents.

11. An apparatus for reconstituting digital watermark data embedded in digital data contents, said method comprising:

a reading means for reconstituting the digital watermark data embedded in the digital data contents from data sequence extracted from the data contents before being reconstituted by using a detection method based on binary distribution function F(x) in statistics.

12. An apparatus as claimed in claim 10, said reading means comprising:

means for obtaining a binary distribution function F(x) from an appearance probability of each bit of 1 bit sequence extracted at random from digital data contents and a repeating number t of each bit of digital watermark data;

means for extracting digital watermark data sequence with respect to each bit value of the digital watermark data;

means for obtaining the appearance probability of the extracted digital watermark sequence from said obtained binary distribution function F(x);

means for judging whether the obtained appearance probability or 1- appearance probability is more than a threshold value of reliability; and

means for reconstituting the digital watermark data if the data is judged to be more than said threshold value.

13. There apparatus as claimed in claim 12, comprising:

means for obtaining reliability regarding said reconstituted data bit from the appearance probability and outputting the same together with the reconstituted data bit.

14. The apparatus as claimed in claim 11, if a data sequence which is embedded as said digital watermark data is modulated by a pseudo-random sequence,

setting a reliability threshold value α of the extracted digital watermark data,

obtaining a binary distribution function F(x)from an appearance probability q of each bit of 1 bit sequence extracted at random from digital data contents and a repeating number t of each bit of digital watermark data;

said apparatus further comprising:

means for demodulating said modulated data sequence by said pseudo-random sequence before the digital watermark data is reconstructed, wherein said reading means comprising,

means for obtaining the number k_i of bit 1 or 0 included in ith digital watermark sequence n_i,

assigning 1/2 to said probability q;

obtaining a maximum number x_0 which satisfies $0 \leqq F(x=x_0) \leqq 1-\alpha$ and a minimum number x_1 which satisfies $\alpha \leqq F(x=x_1) \leqq 1$;

obtaining the number k_i of '1' or '0' included in said ith digital watermark sequence.

15. The apparatus as claimed in claim 1, further comprising:

means for setting a reliability threshold value α of the extracted digital watermark data;

means for obtaining a binary distribution function $F(x)$ from an appearance probability q of each bit of 1 bit sequence extracted at random from digital data contents and a repeating number t of each bit of digital watermark data;

means for obtaining probability whether the watermark sequence is watermark data by using the binary distribution function $F(x)$;

judging means for searching whether the probability is above the threshold value α;

means for reconstituting digital watermark data by performing majority decision processing for said watermark sequence read from the data contents when the judging means judges the probability is above the threshold value α; and

means for judging that there is no digital watermark or the presence is unknown if said probability is not above the threshold value α.

16. The apparatus as claimed in claim 15, further comprising:

a means for outputting said probability as reliability of said reconstituted digital watermark data.

17. The apparatus as claimed in claims 11, 13, 15, 16, a data sequence which is embedded as said digital watermark data is modulated by a pseudo-random sequence, further comprising:

means for demodulating said modulated data contents by said pseudo-random sequence before the digital watermark data is reconstructed.

18. The apparatus as claimed in claim 11,

wherein:

the reliability threshold value α of the extracted digital watermark data is set;

the binary distribution function F(x) from an appearance probability q of each bit of 1 bit sequence extracted at random from digital data contents and a repeating number t of each bit of digital watermark data are obtained;

the data sequence embedded as said digital watermark data is modulated by a pseudo-random sequence,

the watermark apparatus includes a means for demodulating said modulated data contents by said pseudo-random sequence before the digital watermark data is reconstructed; and

said extracting means includes:

a judging means which assigns 1/2 to said probability q, and by comparing a number x_0 which satisfies $0 \leqq F(x=x_0) \leqq 1-\alpha$ and a number x_1 which satisfies $\alpha \leqq F(x=x_1) \leqq 1$ with the average value of bias from the center of the binary distribution of the watermark, judges the average value is equal to or less than the average value or equal to or more than the average value;

means for judging whether the average value is equal to or less than the average value or equal to or

more than the average value;

means for reconstituting digital watermark data by performing majority decision processing for said watermark sequence read from the data contents when the judging means judges the average value is equal to or less than the average value or equal to or more than the average value; and

means for judging that there is no digital watermark or the presence is unknown when the judging means judges the average value is not equal to or less than the average value or not equal to or more than the average value.

19. The apparatus as claimed in claim 17, further comprising means for outputting said average value of the bias from the center of the binary distribution of the watermark sequence as reliability of said reconstituted digital watermark data in case the digital watermark data is reconstructed.

20. A memory medium which stores a program for reconstituting digital watermark embedded in data contents, comprising:

a performing means for reconstituting the digital watermark data embedded in the digital data contents

from data sequence extracted from the data contents before being reconstituted by using a detection method based on binary distribution function F(x) in statistics.

21. A memory medium as claimed in claim 20, wherein a reliability threshold value α of the extracted digital watermark data is predetermined,

said reconstituting process of digital watermark further comprising:

processing means for obtaining a binary distribution function F(x) from an appearance probability q of each bit 1 or 0 bit sequence extracted at random from digital data contents and a repeating number of each bit of digital watermark data, wherein the binary distribution function F(x) represents the probability of x numbers of 1 or 0 in said bit sequence;

processing means for extracting the data watermark sequence n_i from each bit of the digital watermark data from target area for embedding watermarks from the information contents;

processing means for calculating numbers K_i of bit 1 or bit 0 included in the extracted data watermark sequence n_i;

processing means for reconstituting the ith digital watermark data W_i to be 1 or 0 if $F(K_i)>\alpha$;

processing means for reconstituting the digital watermark data W_i to be 0 or 1 if $1-F(K_i)>\alpha$;

processing means for judging that there is no digital watermark or the presence is unknown if it neither $F(K_i)>\alpha$ nor $1-F(K_i)>\alpha$.

22. A memory medium as claimed in claim 21, further comprising:

means for outputting $F(k_i)$ as its reliability when the reconstituted digital watermark data W_i is 1;

processing means made by a computer for outputting $1-F(K_i)$ as its reliability when the reconstituted digital watermark data W_i is 0.

23. A memory medium as claimed in claim 20, wherein a reliability threshold value α of the extracted digital watermark data is predetermined,

said reconstituting process of digital watermark further comprising:

processing means for obtaining a binary distribution function F(x) from an appearance probability q of each bit 1 or 0 bit sequence extracted at random from digital data contents and a repeating number of each bit of digital watermark data, wherein the binary distribution function F(x) represents the

probability of x numbers of 1 or 0 in said bit sequence;

processing means for obtaining probability which represents whether the watermark sequence is the watermark data or not by using the binary distribution function F(x);

processing means for detecting whether said probability is above the threshold value α or not;

processing means for reconstituting digital watermark data by performing majority decision processing for said watermark sequence read from the data contents when the judging means judges the probability is above the threshold value α; and

means for judging that there is no digital watermark or the presence is unknown if said probability is not above the threshold value α.

24. A memory medium as claimed in claim 23, further comprising:

processing means made by said computer for outputting said probability as its reliability of the reconstituted digital watermark data.

25. A memory medium as claimed in any of the claims 20 to 24, wherein said data contents is a modulation of actually embedded digital sequence as

digital watermark data by pseudo-random sequence, the memory medium further comprising:

processing means made by said computer for demodulating said modulated data contents by pseudo-random sequence before the digital watermark data is reconstructed.

26. A memory medium as claimed in claim 20,

wherein said data contents is a modulation of actually embedded digital sequence as digital watermark data by pseudo-random sequence,

wherein a reliability threshold value α of the extracted digital watermark data is predetermined,

wherein a binary distribution function F(x) is obtained from an appearance probability q of each bit 1 or 0 bit sequence extracted at random from digital data contents and a repeating number t of each bit of digital watermark data, wherein the binary distribution function F(x) represents the probability of x numbers of 1 or 0 in said bit sequence,

processing means for reconstituting said digital watermark data comprising:

processing means for demodulating said modulated data contents by pseudo-random sequence before the digital watermark data is reconstructed;

processing means for obtaining numbers K_i of bit 1 or bit 0 included in ith digital watermark data W_i;

processing means for restructuring the ith digital watermark data W_i to 0 or 1 if K_i, which assigns 1/2 to an appearance probability q, is equal to or less than the largest value of x_0 which satisfies $0 \leqq F(x=x_0) \leqq 1-\alpha$; and

processing means for restructuring the digital watermark data W_i to 1 or 0 if K_i, which assigns 1/2 to an appearance probability q, is equal to or more than the smallest value of x_1 which satisfies $\alpha \leqq F(x=x_1) \leqq 1$.

27. A memory medium as claimed in claim 20,

wherein said data contents is a modulation of actually embedded digital sequence as digital watermark data by pseudo-random sequence,

wherein a reliability threshold value α of the extracted digital watermark data is predetermined,

wherein a binary distribution function F(x) is obtained from an appearance probability q of each bit 1 or 0 bit sequence extracted at random from digital data contents and a repeating number t of each bit of digital watermark data, wherein the binary distribution function F(x) represents the probability of x numbers of 1 or 0 in said bit sequence,

wherein x_0 or x_1,which assign 1/2 to an appearance probability q, are obtained which satisfies $0 \leqq F(x=x_0) \leqq 1-\alpha$ or $\alpha \leqq F(x=x_1) \leqq 1, \alpha$,

the memory medium further comprising:

processing means for demodulating said modulated data contents by pseudo-random sequence before the digital watermark data is reconstructed;

processing means for judging the average value of the bias from the center of the binary distribution of the watermark sequence is equal to or less than the average value x_0 or equal to or more than the average value x_1;

processing means for reconstituting the digital watermark data;

processing means for reconstituting digital watermark data by performing majority decision processing for said watermark sequence read from the data contents when the judging means judges the probability is equal to or less the average value x_0 or equal to or more than the average value x_1; and

processing means for judging that there is no digital watermark or the presence is unknown if said probability is not equal to or less the average value x_0 or not equal to or more than the average value x_1.

[Detailed Description of the Invention]

[Field of the Invention]

The present invention generally relates to a digital watermarking technique. More particularly, the present invention relates to a digital watermarking technique for embedding or reading digital watermark data in digital data contents which represent an image or a sound. In addition, the present invention relates to a technique for statistical processing of read watermark data in a system using the digital watermarking technique.

[Prior Art]

Conventional techniques are proposed in Japanese patent applications No.8-305370, No.8-338769, No.9-9812, No.9-14388, No.9-109924, No.9-197003, No.9-218467 and No.10-33239. The digital watermark method is also called data hiding, finger printing steganography, image/sound deep encryption and the like. As for a digital watermarking system, accuracy for determining the presence or absence of embedded data is important. In addition, reliability of embedded data is important. The digital watermarking system generally has a mechanism for reconstituting correct digital watermark data even when sub-data embedded in the data contents is corrupted to a certain extent, since the digital

watermarking system assumes various processing on the watermarked data contents. However, under present circumstances, it is impossible for the system to evaluate validity of reconstituted digital watermark data quantitatively. Therefore, the system does not have enough reliability.

[Object of the Invention]

It is an object of the present invention to evaluate quantitatively probabilities of cases that data contents which do not contain digital watermark data are wrongly judged as containing digital watermark data, and incorrect digital watermark data is read from watermarked digital data contents.

[Means to Solve the Problems]

Digital watermark processing is comprised of a pair of digital watermark embedding/digital watermark extraction. In digital watermark embedding processing, digital watermark embedding area $B \in A$ is selected from the digital watermark target area in data contents by means of secret key information so that data in area B is changed by an inherent rule. In digital watermark extraction processing, data of the digital watermark embedding area B is interpreted and the digital watermark data is reconstituted. The present invention judges the probability of occurrence of the digital

watermark data read from the watermark embedding area B by means of the correct secret key data, based on a binary distribution in statistics of digital watermark data read by means of any secret key data regardless of true or false from A which is an overall digital watermark target area, by means of the digital watermark algorism which is to be an applied target of the invention, in the data contents embedded with digital watermark.

Effect

The present invention, in an digital watermark technique, can evaluate the credibility of the watermark data read from the data contents, can judge whether the data contents have the digital watermark or not, and can suppress the probability of reading incorrect digital watermark in an certain value from data contents which the digital watermark is included.

[Embodiments of the Invention]

Embodiment 1

Before explaining embodiments of the present invention, definition of some words will be given. "Digital watermark data sequence" represents a data sequence read from the digital data contents before being reconstituted. "Digital watermark data" represents significant data for system operation, which data needs

to be embedded in the digital data contents, or, data obtained by reconstituting the digital watermark sequence. "Reliability α of digital watermark" is an index representing validity of read digital watermark data. That is, it represents a probability that the read digital watermark data matches with the actual embedded digital watermark data. Conversely, a probability of reading digital watermark data from an image without digital watermark data or reading erroneous digital watermark data can be represented as $2(1-\alpha)$.

Similarly, "embedded sequence" represents data to be actually embedded. The embedded sequence includes sequence of embedded data which is modulated, extended or repeated. In addition, "read" may be replaced with "extract" in some cases.

Fig.1 shows a digital watermarking system of the present invention. In the system shown in Fig.1, digital watermark data 101 is embedded in digital data contents 103 by a digital watermark embedding apparatus 102, then, converted into watermarked digital data contents 104.

The watermarked digital data contents 104 are degraded to watermarked digital data contents 105 by compression or image processing while the watermarked

digital data contents 104 are distributed by wireless or wire communication or by a packaged medium.

A digital watermark reading apparatus 107 reads a watermark sequence from the degraded watermarked digital data contents 105, and reconstitutes digital watermark data 108.

Fig.2 is a block diagram of the watermark reading apparatus 107. The digital watermark data reconstitution apparatus 108 provided in the watermark reading apparatus 107 obtains the probability q that bit 1 is read when any 1 bit watermark sequence is read from a whole watermark area beforehand by using the watermark reading apparatus 107.

Specifically, assuming a 1 bit watermark sequence reading part 501, the part 501 reads the watermark sequence 1 bit by 1 bit from all elements of the whole watermark area (a broken line L1), and calculates the ratio of the number of bit 1 to the number of all trials.

In the embodiment, the reading probability of bit 1 and the number of bit 1 are obtained. However, it is possible that the reading probability of bit 0 and the number of bit 0 are obtained. Basically, there is no difference between the former and the latter. The difference is only on implementation.

Accordingly, the probabilities of detecting bit 0 and 1 when reading 1 bit at random in the watermark area by using the digital watermarking algorithm is calculated to be 1-q and q respectively.

The n bit watermark sequence reading part 502 reads the digital watermark data sequence from the watermarked digital data contents for the number of total times of embedding digital watermark data.

Here, digital watermark data is defined as b_0, b_1, …, b_{m-1}, $b_i \in \{0, 1\}$, $i < m$ (m bit length), the repeating number of embedding ith bit of the digital watermark data in the digital data contents is defined as n_i, the read watermark sequence is defined as $b'_{0,0}$, $b'_{0,1}$, …$b'_{0,n0-1}$, $b'_{1,0}$, $b'_{1,1}$, … $b'_{1,n1-1}$,… ,$b'_{m-1,0}$, $b'_{m-1,1}$, … $b'_{m-1,nm-1-1}$ $b_{i,j} \in \{0, 1\}$ ($\sum_{r=0}^{m-1} n_r$ bit length).

The data reconstitution apparatus 108 receives a subsequence of the digital watermark data sequence one after another from a subsequence corresponding to 0th digital watermark data to a subsequence corresponding to (m-1)th digital watermark data (a solid line L2).

Next, the method for reconstituting ith bit of the digital watermark data will be described concretely.

When n_i bits of digital watermark data sequence is read at random from the watermark area, the probability P(x=k) of k '1' bits appearing in the n_i bit

sequence is represented by the binary distribution density function

$$P(x=k)=n_iC_kq^k\cdot(1-q)^{ni-k} \quad (1)$$

and the distribution function of that, F(x), is

$$F(x)=\sum_{k=0}^{x} n_iC_kq^k\cdot(1-q)^{ni-k} \ (0\leqq x\leqq n_i). \quad (2)$$

Here, n_iC_k is the number of combinations when selecting k out of n_i.

Setting a reliability threshold value α ($1/2<\alpha\leqq 1$) of the digital watermark data, the number of bit 1 included in a subsequence $b'_{i,0}$, $b'_{i,1}$, ...$b'_{i,ni-1}$ corresponding to ith digital watermark data is calculated by

$$k_i = \sum_{r=0}^{ni-1} b'_{i,r} .$$

Then, digital watermark data is determined in the following way by using the formula (2):

$$b_i = \begin{cases} 0 & \text{when } 0\leqq F(k_i)\leqq 1-\alpha \\ 1 & \text{when } \alpha\leqq F(k_i)\leqq 1 \\ \text{unknown or not present} & \text{when } 1-\alpha< F(k_i)<\alpha \end{cases} \quad (3)$$

Viewing from a different angle, when determining by the number of bit 1 included in the watermark sequence n_i, if the largest integer x_0 that satisfies $0\leqq F(x = x_0) \leqq 1-\alpha$ and the smallest integer x_1 that satisfies $\alpha\leqq F(x = x_1) \leqq 1$ are assumed to be

threshold values, the digital watermark data is judged as shown in Fig.3 such that if the number of 1 in n_i is equal to or smaller than x_0 , the digital watermark data is 0, and that if the number of 1 is equal to or larger than x_1, the digital watermark data is 1.

The horizontal axis of Fig.3 represents the number of bit 1 included in the watermark sequence, and the vertical axis represents frequency of the corresponding number. As for unwatermarked digital data contents, the frequency that bit 1 appears in a bit sequence read at random from the digital data contents becomes binary distribution. Thus, the peak of the frequency is at the half point of the number of bits. On the other hand, as for watermarked contents, in the subsequence n_i in which bit 0 is embedded as digital watermark data, the frequency of bit 1 is 0 if there is no degradation and it is a small number which is equal to or smaller than x_0 even if there is degradation. In the subsequence n_i in which bit 1 is embedded as digital watermark data, the frequency of bit 1 is n1 if there is no degradation and it is a large number which is equal to or larger than x_1 even if there is degradation. In this way, the distribution of the frequency of bit 1 or bit 0 in the watermarked sequence is leaning to one side from the center of the binary distribution. The present

invention uses the lean for reconstituting digital watermark data from the read watermark sequence.

Depending on a watermarking system, a following method can be used. That is, reconstituted digital watermark data is obtained by using the bias from the central value of the distribution P(x) of the watermark sequence extracted from digital data contents 105. Next, the probability of appearing the read watermark sequence is calculated by the formula (2). Then, if the reconstituted digital watermark data is 1, $F(k_i)$ can be added to watermark dada as the reliability, and, if the reconstituted digital watermark data is 0, $1-F(k_i)$ can be added. The reliability $F(k_i)$ and $1-F(k_i)$ of the digital watermark data is obtained from the bias of appearance probability of the digital watermark data in the binary distribution of appearance probability of each bit of 1 bit sequence extracted at random from digital data contents.

Fig.4 shows a concept in which the length of the digital watermark data is extended to m bits.

The digital watermark data reconstitution apparatus 108 outputs the reconstituted digital watermark data b_0, b_1, …, b_{m-1} as read digital watermark data 106.

Fig.5 is a flowchart showing the above-

mentioned process. The process will be described in the following with reference to Fig.5.

Watermarked digital data contents 105 and key data which is necessary for reading digital watermark data is input, and a digital watermark data sequence is extracted with respect to each bit value in step 1. Then, a threshold value α of the reliability is set in step 2, and a probability q that bit 1 appears when 1 bit of digital watermark data is read at random from the whole watermark area is obtained in step 3. Then, a binary distribution function F(x) which represents probability that x bits of 1 are included in the bit sequence is obtained from the probability q and the repeating number n_i of each bit of digital watermark data in step 4.

Then, 0 is assigned to i which distinguish a subsequence of the digital watermark data sequence in step 5. Next, the number of bit 1 in the subsequence is obtained as $k_i = \sum_{r=0}^{ni-1} b'_{i,r}$ and the appearance probability $F(k_i)$ is obtained, then it is determined whether $F(k_i)$ is equal to or less than $1-\alpha$ in step 6. If $F(k_i) \leqq 1-\alpha$, the digital watermark data w'_i is reconstituted as 0 in step 7. Then, i is incremented by 1 in step 8, and the process goes back to step 6 if i<m in step 9. If $F(k_i) \leqq 1-\alpha$ is not true in step 6, it is

checked whether $F(k_i) \geqq \alpha$ is true in step 10. If $F(k_i) \geqq \alpha$, the digital watermark data w_i is reconstituted as 1 in step 11, and the process goes to step 8. If $F(k_i) \geqq \alpha$ is not true in step 10, the process ends by determining as there is no watermark or the presence or absence is unknown in step 12. If i is more than n_i in step 9, a reconstituted watermark sequence $\{w'_i\}$ is output. In the above process, the reading process in step 1 can be carried out between step 4 and step 5. In step 6, it is checked whether $1-F(k_i)$ is more than α.

In the first embodiment, it is assumed that there is no bias in the distribution represented by formula (1), that is, $q \cong 1/2$.

When the embedding number n_i of each bit of digital watermark data is adequate for obtaining a statistical characteristic, it becomes $q \cong 1/2$ generally. However, since the value of q depends on characteristics of an watermarking algorithm and digital data contents, q may take a value deviating largely from 1/2 in some rare cases. A method for solving this problem will be described in a second embodiment.

Second Embodiment

In the following, the second embodiment will be described. Fig.6 is a block diagram of a watermarking system of the second embodiment.

The watermark embedding apparatus 102 embeds digital watermark data 101 in digital data contents 103. At the time, when embedding each bit value n_i times repeatedly, watermark sequence is modulated and embedded in the digital data contents 103. The modulation is carried out by a pseudo-random sequence generator (A) 501 which is provided in the watermark embedding apparatus 102.

For example, when assuming the embedding sequence as $b_{0,0}$, $b_{0,1}$, $\ldots b_{0,n0-1}$, $b_{1,0}$, $b_{1,1}$, $\ldots$ $b_{1,n1-1}$, $\ldots$, $b_{m-1,0}$, $b_{m-1,1}$, $\ldots b_{m-1,nm-1-1}$ $b_{i,j} \in \{0, 1\}$, and the pseudo-random sequence as $r_{i,0}$, $r_{i,1}$, $\ldots r_{i,ni-1}$ $b_{i,j} \in \{0, 1\}$, the embedding sequence is modulated to

$m_{i,0}$, $m_{i,1}$, $\ldots m_{i,ni-1}$

$m_{i,j} = b_{i,j}$ (+) $r_{i,j}$

by the pseudo-random sequence. A(+)B represents XOR of A and B.

According to the above-mentioned process, the same pseudo-random sequence is necessary for digital watermark data reading.

For example, if 1 bit watermark sequence is read by using an M-sequence as the pseudo-random sequence, it becomes $q \cong 1/2$. Therefore, the present invention can be applicable without depending on the watermarking algorithm and digital data contents.

When digital watermark data reading, demodulation is carried out as $b'_{i,j} = m_{i,j} (+) r_{i,j}$ by using a pseudo-random sequence generator (B) 502 which is provided in the watermark reading apparatus 106.

Here, the pseudo-random sequence generator (A) 501 and the pseudo-random sequence generator (B) 502 needs to be implemented such that both of the generators generate the same pseudo-random sequence.

Watermark data is reconstituted with the method of the first embodiment from the watermark sequence $b'_{0,0}$, $b'_{0,1}$, $\ldots b'_{0,n0-1}$, $b'_{1,0}$, $b'_{1,1}$, $\ldots$ $b'_{1,n1-1}$, $\ldots$, $b'_{m-1,0}$, $b'_{m-1,1}$, $\ldots b'_{m-1,nm-1-1}$ $b_{i,j} \in \{0, 1\}$ obtained by the demodulation.

Since it is considered that the appearance probability q of bit 1 in the watermark sequence can be approximated by the binary distribution regardless of the presence or absence of modulation, there is no influence on the distribution of the density function (1) due to the modulation shown in this embodiment.

In addition, q=1/2 can be assumed in implementation, that is, no process is necessary for obtaining q. Therefore, the amount of processing that is required for watermark reconstitution thus becomes the same as that for majority decision processing. Thus, the reconstitution process becomes faster.

Third Embodiment

In the following, a third embodiment will be described. In the third embodiment, an example will be described showing concrete values on the basis of the first embodiment and the second embodiment. In this embodiment, it is assumed that digital watermark data is 1 bit, the repeating number n of embedding is 127 and the probability q that bit 1 is read when reading 1 bit watermark sequence at random from the whole watermark area is 1/2. If the threshold value α is 0.99999 (which means 99.999%), x_0 in Fig.21 is 36 and x_1 is 90. That is to say, according to the present invention, under the above-mentioned condition, digital watermark data is judged as bit 0 if the number of '1' appeared in the watermark sequence (n bits) is equal to or less than 36, and it is judged as bit 1 if the number of '1' appeared in the watermark sequence (n bits) is equal to or more than 90, and it is judged that there is no watermark data or the presence or absence is unknown in other cases. If it is judged that there is digital watermark data, the correctness of more than 99.999% can be ensured.

Fourth Embodiment

A fourth embodiment will be described in the following. According to the embodiment shown in Fig.5,

as is understood from the above-mentioned procedure, if even the reliability of only 1 bit is not obtained, that is, if $F(k_i)$ or $1-F(k_i)$ is less than α, the reconstitution of the digital watermark data becomes impossible because it is judged that there is no digital watermark data or the presence or absence is unknown. The fourth embodiment solves the problem. In this case, it is assumed that the repeating number of embedding each bit of digital watermark data is the same value n.

The method for reconstituting digital watermark data w_0, w_1, …, w_{m-1} from watermark sequence $b'_{0,0}$, $b'_{0,1}$, …$b'_{0,n0-1}$, $b'_{1,0}$, $b'_{1,1}$, … $b'_{1,n-1}$,… ,$b'_{m,0}$, $b'_{m,1}$, …$b'_{m,n-1}$ which is read from digital data contents will be described in the following with reference to Fig.7.

The watermark sequence is read with respect to each bit value from the digital data contents and key data necessary for digital watermark data reading in step 1.

The threshold value α $(1/2<\alpha\leqq 1)$ of the reliability is set in step 2. For example, if the reliability of read digital watermark data needs to be equal to or more than 99%, it is set as $\alpha=0.99$.

The probability q of bit '1' when 1 bit of the watermark sequence is read at random from the whole watermark area of watermarked digital data contents is

obtained beforehand in step 3. The appearance probabilities of bits '0' and '1' are calculated as 1-q and q respectively.

The probability that x bits of '1' are included in the watermark sequence of each bit data of digital watermark data are obtained as

$$F(x)=\sum_{j=0}^{x} {}_nC_j q^j \cdot (1-q)^{n-j}$$

by using the binary distribution function in step 4.

It is checked in step 5 that the probability that n bit digital watermark data sequence is digital watermark data exceeds the threshold value α. Specifically, it is checked whether the following formula (4) is satisfied.

$$F\left(\frac{\sum_{i=0}^{m-1}\left|\sum_{j=0}^{n-1} b'_{i,j}-\frac{n}{2}\right|}{m}+\frac{n}{2}\right)\geq\alpha \qquad (4)$$

Here, |a| represents the absolute value of a. $\sum_{j=0}^{n-1} b_{i,j}$-n/2 represents the bias from the center of the binary distribution of the number of bit '1' in the n bit watermark sequence. $\sum_{i=0}^{m-1}$ of $\sum_{j=0}^{n-1} b_{i,j}$-n/2 divided by m represents the average for the m bits of the whole digital watermark data. n/2 represents the center of the binary distribution.

If the formula (4) is true, it is judged that

there is digital watermark data. Thus, in each n bit watermark sequence of m digital watermark data sequences, digital watermark data is reconstituted by a majority decision processing.

Specifically, if it is judged that there is digital watermark data, digital watermark data is reconstituted in the following way in step 6.

For all i ($0 \leqq i < m$),

when $\sum_{j=0}^{n-1} b_{i,j} < n/2$: $w'_i = 0$,

when $\sum_{j=0}^{n-1} b_{i,j} > n/2$: $w'_i = 1$.

This process is carried out by steps 6-1 - 6-7 in Fig.7.

If $$F\left(\frac{\sum_{i=0}^{m-1}\left|\sum_{j=0}^{n-1} b_{i,j} - \frac{n}{2}\right|}{m} + \frac{n}{2}\right) < \alpha ,$$

it is judged that there is no watermark data or the presence or absence is unknown. A following formula (5) can be used instead of the formula (4).

$$F\left(\frac{n}{2} - \frac{\sum_{i=0}^{m-1}\left|\sum_{j=0}^{n-1} b'_{i,j} - \frac{n}{2}\right|}{m}\right) \leq 1-\alpha \qquad (5)$$

If the formula (5) is not true, it is judged that there is no watermark data or the presence or absence unknown.

According to the fourth embodiment,

statistical processing for whole watermark sequence is carried out so as to judge the presence or absence of watermark by using the formula (4) or the formula (5). If it is judged that there is digital watermark data, the reconstitution is carried out by the majority decision processing. Therefore, even if there is one bit of low reliability, digital watermark data can be reconstituted.

In Fig.7, the step 1 can be carried out between the steps 4 and 5.

The fourth embodiment may use the pseudo-random sequence which is described in the second embodiment. Specifically, watermark embedding is carried out by modulating digital watermark data sequence with the pseudo-random sequence. When reconstituting, the read digital watermark data sequence is demodulated by the pseudo-random sequence, then the judgment by the formula (4) is performed . If the result is more than α and there is digital watermark data, the reconstitution process of the majority decision is performed on the demodulated sequence, which is the same process as the step 6 of the fifth embodiment. The whole process is shown in Fig.26, adding the same reference symbol to the corresponding part shown in Fig.7. In the example, the pseudo-random

sequence $\{r_{i,j}\}$ is generated from key data 'Key' and the process goes to step 2 in step 8. Next to step 4, watermark sequence is demodulated with the pseudo-random sequence $\{r_{i,j}\}$ in step 9. The watermark bit $b'_{i,j}$ in the formula (4) in step 5 is a bit which is demodulated in step 9. Also, the majority decision processing in step 6 is performed on $b'_{i,j}$.

Fifth Embodiment

In the following, a fifth embodiment will be described.

Since digital watermark data is dispersed by the pseudo-random sequence, when q is approximated to 1/2, the presence or the absence of watermark data in the watermark sequence can be judged as follows.

The probability that x bits of '1' (a number x of '1' bits) are included in the n bit watermark sequence which constitutes each bit of digital watermark data is represented as

$$F(x) = \sum_{j=0}^{x} {}_nC_j (1/2^n)$$

by using the binary distribution function. Accordingly, by obtaining the smallest integer x_1 which satisfies $F(x=x_1) \geqq \alpha$, the demodulated sequence of the step 5 in the tenth embodiment can be judged with the following formula (6).

$$\frac{\sum_{i=0}^{m-1}\left|\sum_{j=0}^{n-1} b'_{i,j} - \frac{n}{2}\right|}{m} + \frac{n}{2} \geq x_1 \quad \cdots (6)$$

In this case, the amount of processing can be reduced to the same level as that of the majority decision processing.

The judgment is equivalent to a judgment for judging whether the average of the bias from the center n/2 of the binary distribution of the watermark sequence is equal to or more than x_1.

If the formula (6) is true and it is judged that there is digital watermark data, the majority decision process is performed on the watermark sequence which is demodulated by the pseudo-random sequence in the following way.

For all i ($0 \leqq i < m$),

when $\sum_{j=0}^{n-1} b'_{i,j} < n/2$: $w'_i = 0$,

when $\sum_{j=0}^{n-1} b'_{i,j} > n/2$: $w'_i = 1$

Then, the digital watermark data is reconstituted.

If $$\frac{\sum_{i=0}^{m-1}\left|\sum_{j=0}^{n-1} b'_{i,j} - \frac{n}{2}\right|}{m} + \frac{n}{2} < x_1,$$

it is judged that there is no watermark data or the presence or absence is unknown.

In the above process, it is possible to use

the maximum integer x_0 which satisfies $F(x=x_0) \leqq 1-\alpha$ instead of the minimum integer x_1 which satisfies $F(x=x_1) \geqq \alpha$. In this case, a formula for judging the presence or absence of watermark is shown below as a formula (7).

$$\frac{n}{2} - \frac{\sum_{i=0}^{m-1} \left| \sum_{j=0}^{n-1} b'_{i,j} - \frac{n}{2} \right|}{m} \leq x_0 \quad \cdots (7)$$

If the left part of the formula is more than x_0, it is judged that there is no watermark data or the presence or absence is unknown.

Sixth Embodiment

In the following, a sixth embodiment of the present invention will be described.

When it is judged that there is digital watermark data by the formula (4), the digital watermark data is reconstituted by the above-mentioned majority decision process. At the same time, the reliability of the reconstituted watermark sequence as a whole is calculated as

$$F\left(\frac{\sum_{i=0}^{m-1} \left| \sum_{j=0}^{n-1} b'_{i,j} - \frac{n}{2} \right|}{m} + \frac{n}{2} \right)$$

and it is output.

Similarly, when it is judged that there is digital watermark data by the formula (5) and the digital watermark data is reconstituted, the reliability of the reconstituted digital watermark data sequence as a whole is calculated as

$$F\left(\frac{n}{2}-\frac{\sum_{i=0}^{m-1}\left|\sum_{j=0}^{n-1}b'_{i,j}-\frac{n}{2}\right|}{m}\right)$$

and it is output.

When it is judged that there is digital watermark data by the formula (6), the digital watermark data is reconstituted by the above-mentioned majority decision process. At the same time, the reliability of the reconstituted watermark sequence as a whole is calculated as

$$F\left(\frac{\sum_{i=0}^{m-1}\left|\sum_{j=0}^{n-1}b'_{i,j}-\frac{n}{2}\right|}{m}+\frac{n}{2}\right)$$

and it is output.

Similarly, when it is judged that there is digital watermark data by the formula (7), the reliability of the digital watermark data as a whole is calculated as

$$F\left(\frac{n}{2}-\frac{\sum_{i=0}^{m-1}\left|\sum_{j=0}^{n-1}b'_{i,j}-\frac{n}{2}\right|}{m}\right)$$

and it is output.

In the above-mentioned first - sixth embodiments, the reading probability of bit 1 and the number of bit 1 are obtained. However, it is possible that the reading probability of bit 0 and the number of bit 0 are obtained. Basically, there is no difference between the former and the latter. The difference is only on implementation.

In the following, examples of experiments will be shown. In the following experiments, an image of "lena" which has 128×128 pixels is used as a test image, and the threshold value α of the reliability is assumed to be 0.999999.

First Experiment

In this experiment, 1 bit digital watermark data '1' was embedded 127 times repeatedly using key data '50,000', and the watermark sequence was read with various key data. Fig.9 shows the number of bit '1' in the read watermark sequence corresponding to the key data. In Fig.9, the vertical axis shows the number of bit '1' in the read watermark sequence, and the

horizontal axis shows the key data value. In this experiment, the appearance frequency of bit '1' in the watermark area A was q=0.492247.

When correct key data (50,000) is used, it is judged that digital watermark data is '1' with 99.9999% correctness since the number of bit '1' is more than the threshold value x_1 for judging the presence of watermark. When incorrect key data is used, it is judged that there is no watermark data or the presence or absence is unknown.

Second Experiment

In the second experiment, a watermark sequence which was modulated with a 7 stage M-sequence (initial state is 64) was embedded, and a similar experiment as the first experiment was carried out with various key data and M-sequences of various initial states. The result is shown in Fig.10. By carrying out the modulation, the value of q becomes 0.500000 from 0.492247, and the variance becomes 31.718777 from 31.008265. Thus, the values are almost not changed from those of the first experiment. It is only when correct key data and correct pseudo-random sequence are used that digital watermark data can be read. In addition, when the watermark sequence is embedded in half data of the watermark area A, q=0.741547 with the modulation and

q=0.499768 without the modulation.

[Advantages of the Invention]

The effects of the present invention corresponding to the second object is as follows.

(1) There are following effects by judging digital watermark data on the basis of the binary distribution in statistics:

- The probabilities of following cases can be evaluated quantitatively. The cases are: digital data contents which do not contain digital watermark data are wrongly judged as containing digital watermark data, and incorrect digital watermark data is read from watermarked digital data contents. In addition, the probability can be suppressed within $2(1-\alpha)$ by using the reliability threshold α of digital watermark data.

(2) There are following effects by modulating digital watermark data by a pseudo-random sequence before embedding the digital watermark data:

- The bias of the probability q of reading bit '1' when 1 bit watermark sequence is read at random from the whole watermark area.

- It becomes difficult to detect the presence or absence of watermark data and the value from the bias of q without the correct key data and the pseudo-random

sequence, the key data being necessary for reading digital watermark data and the pseudo-random sequence being necessary for demodulating read watermark sequence. It can strengthen security which is an important element for the digital watermarking system.

- In an implementation, since it can be assumed to be q=1/2, the amount of processing that is required for watermark reconstitution becomes the same as that for majority decision processing. Thus, the speed of the processing becomes higher.

α is an index which represents a lower limit of the correctness rate of read digital watermark data, and is manageable in the digital watermarking system. Therefor, the method of using α is superior to a conventional method of showing the correctness rate of read digital watermark data to a user.

According to the first embodiment, if there is even one bit of low reliability in digital watermark data $\{w'_i\}$, it is judged that there is no watermark data or the presence or absence is unknown. However, even in the case, according to the eleventh - thirteenth embodiments, the reliability of digital watermark data can be evaluated quantitatively, the probability for reading digital watermark data incorrectly can be suppressed within $2(1-\alpha)$, and the digital watermark

data can be reconstituted. In the tenth - twelfth embodiments, the whole digital watermark data is statistically processed by modifying the formula for judging the presence or absence of digital watermark data, since digital watermark data can be reconstituted in many cases, when watermark sequence $\{b'_{i,j}\}$, $(0 \leqq i<m, 0 \leqq j<n)$ is seen statistically as a whole.

In addition, according to the first embodiment, $F(\sum_{j=0}^{n-1} b'_{i,j})$ needs to be calculated with the distribution function $F(x)$ of the binary distribution for all i to reconstitute digital watermark data from watermark sequence. On the other hand, according to the fourth - sixth embodiments, only one calculation using the distribution function is necessary so that the amount of processing can be reduced.

The present invention becomes more effective in combination with an error correction code. That is, when a part of bits in digital watermark data is intensively corrupted, it is judged that only the part of bits is unknown and other bit data is in high correctness rate. Therefore, correct data can be read by correcting only the corrupted bit data.

[Brief Description of the Drawings]

Fig.1 is a diagram showing an overview of a digital watermarking system;

Fig.2 is a diagram showing an overview of digital watermarking extracted apparatus;

Fig.3 is a diagram showing a judgment of digital watermarking data;

Fig.4 is a diagram showing an overview of digital watermarking data reconstitution;

Fig.5 is a flowchart showing steps of a digital watermark extraction process;

Fig.6 is a diagram showing an overview of the second embodiment of this invention;

Fig.7 is a diagram showing processing steps of the fourth embodiment of this invention;

Fig.8 is a flowchart showing processing steps of the fourth embodiment of this invention in case a target is an embedding of the modulated pseudo-random number watermark data;

Fig.9 is a diagram showing result (no modulation) of the watermark sequence readings;

Fig.10 is a diagram showing result (with modulation) of the watermark sequence readings.

[Name of the Document] Abstract

[Abstract]

[Object] Object of this invention is to quantitatively evaluate the probability of reading the incorrect digital watermark from the data contents which include the digital watermark.

[Solution Means] When n_i bits of digital watermark data sequence is read at random from the watermark area, the probability P(x=k) of k '1' bits appearing in the n_i bit sequence is represented by the binary distribution density function

$$P(x=k)={}_{n_i}C_k q^k \cdot (1-q)^{n_i-k} \quad (1)$$

and the distribution function of that, F(x), is

$$F(x)=\sum_{k=0}^{x} {}_{n_i}C_k q^k \cdot (1-q)^{n_i-k} \ (0 \leqq x \leqq n_i). \quad (2)$$

Here, ${}_{n_i}C_k$ is the number of combinations when selecting k out of n_i.

Setting a reliability threshold value α ($1/2 < \alpha \leqq 1$) of the digital watermark data,

[Selected Figure] Fig. 3

【書類名】図面 [Name of the Document] DRAWING

【図1】FIG. 1

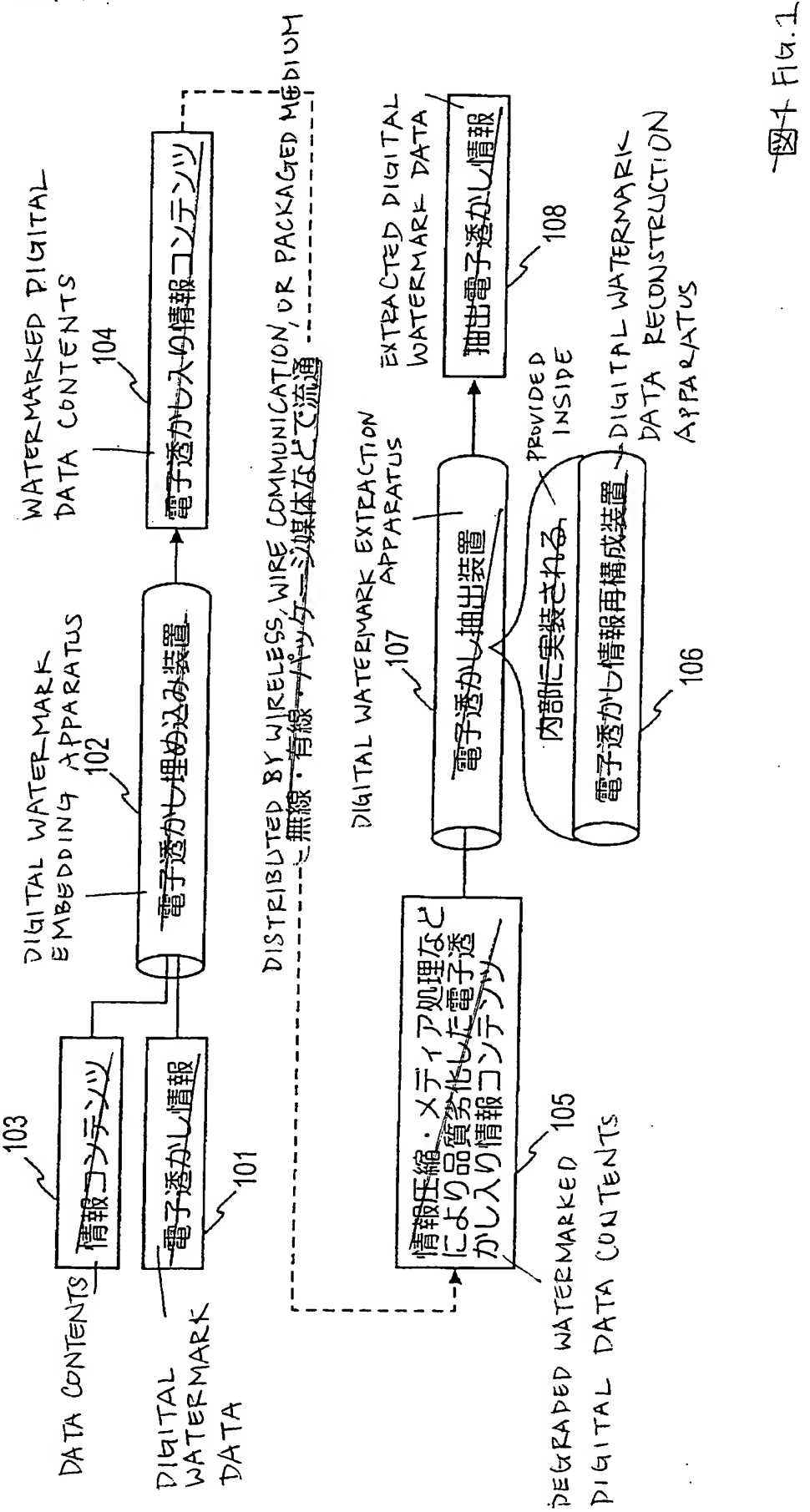


図1 FIG. 1

【図2】 Fig.2

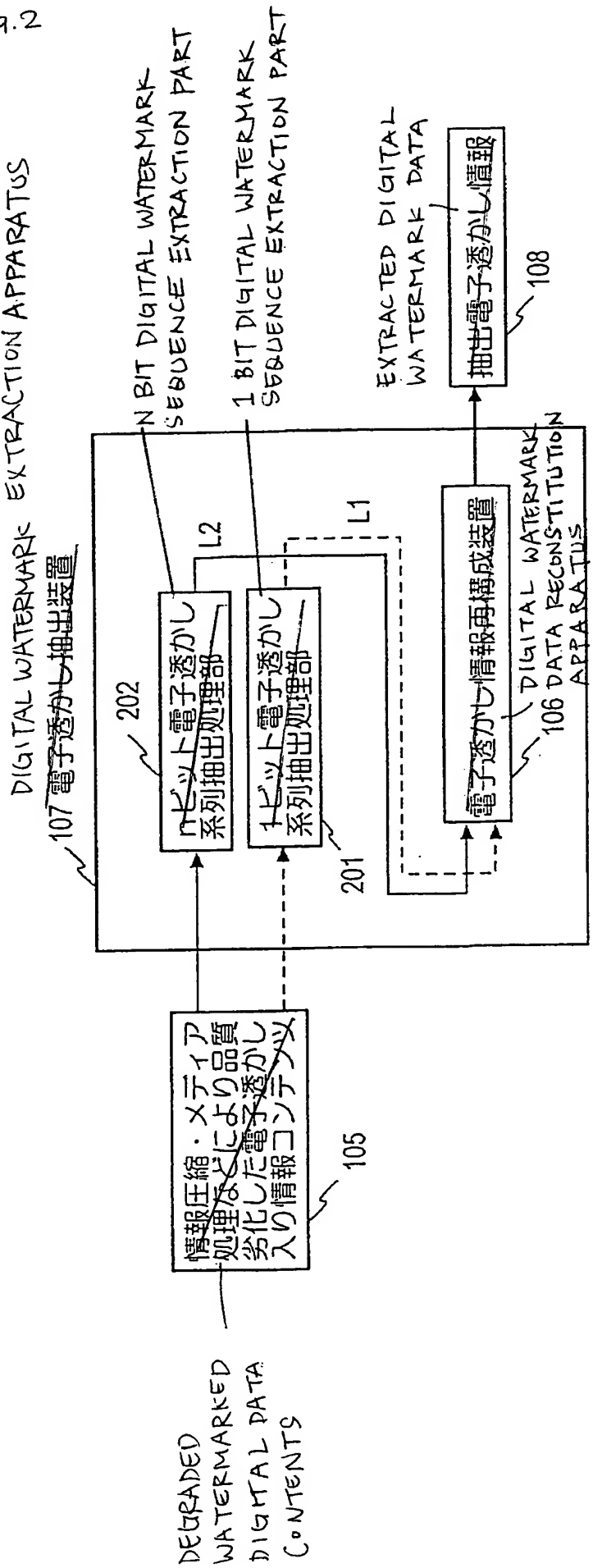


図2 Fig.2

【図3】FIG.3

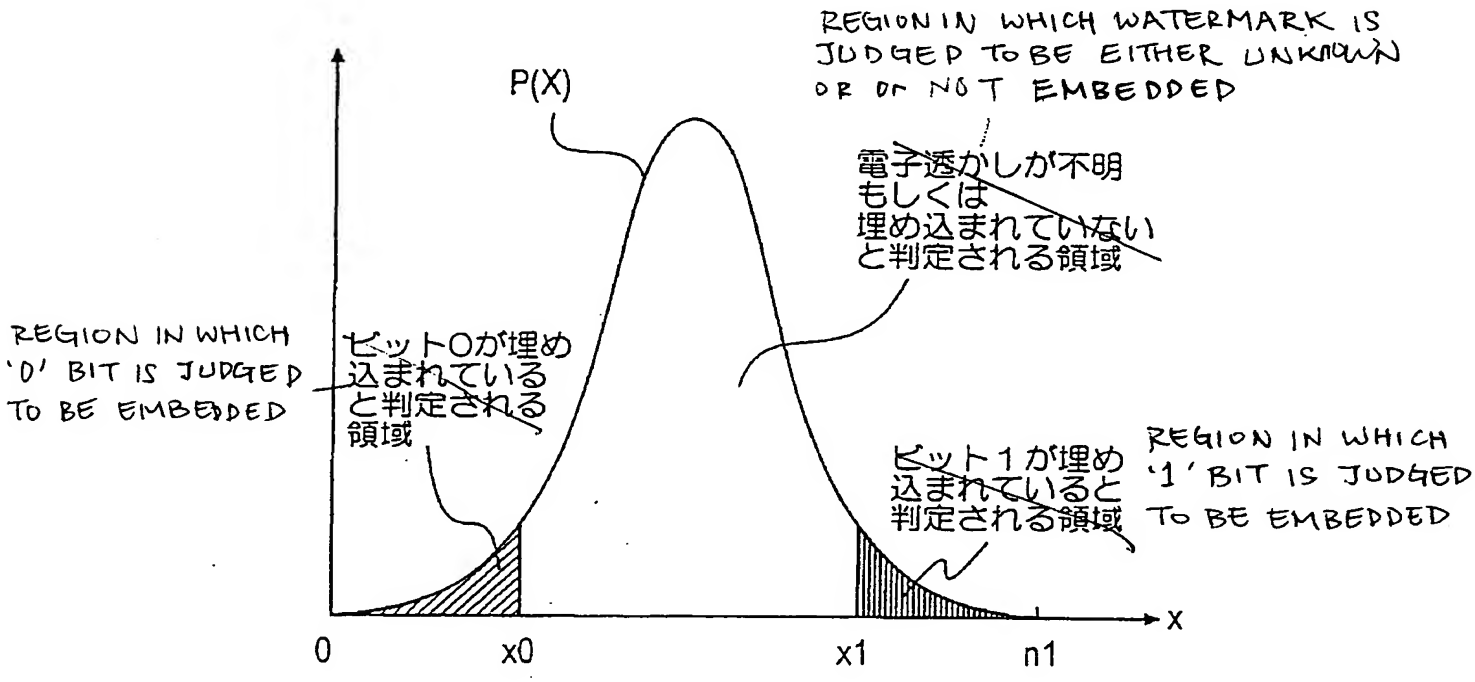


図3
FIG.3

【図4】FIG.4

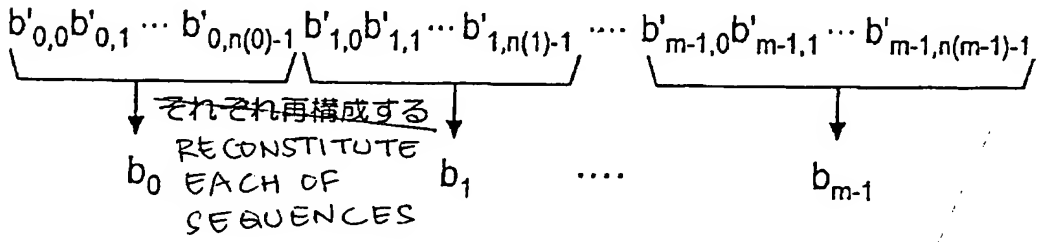


図4
FIG.4

FIG.5 【図5】

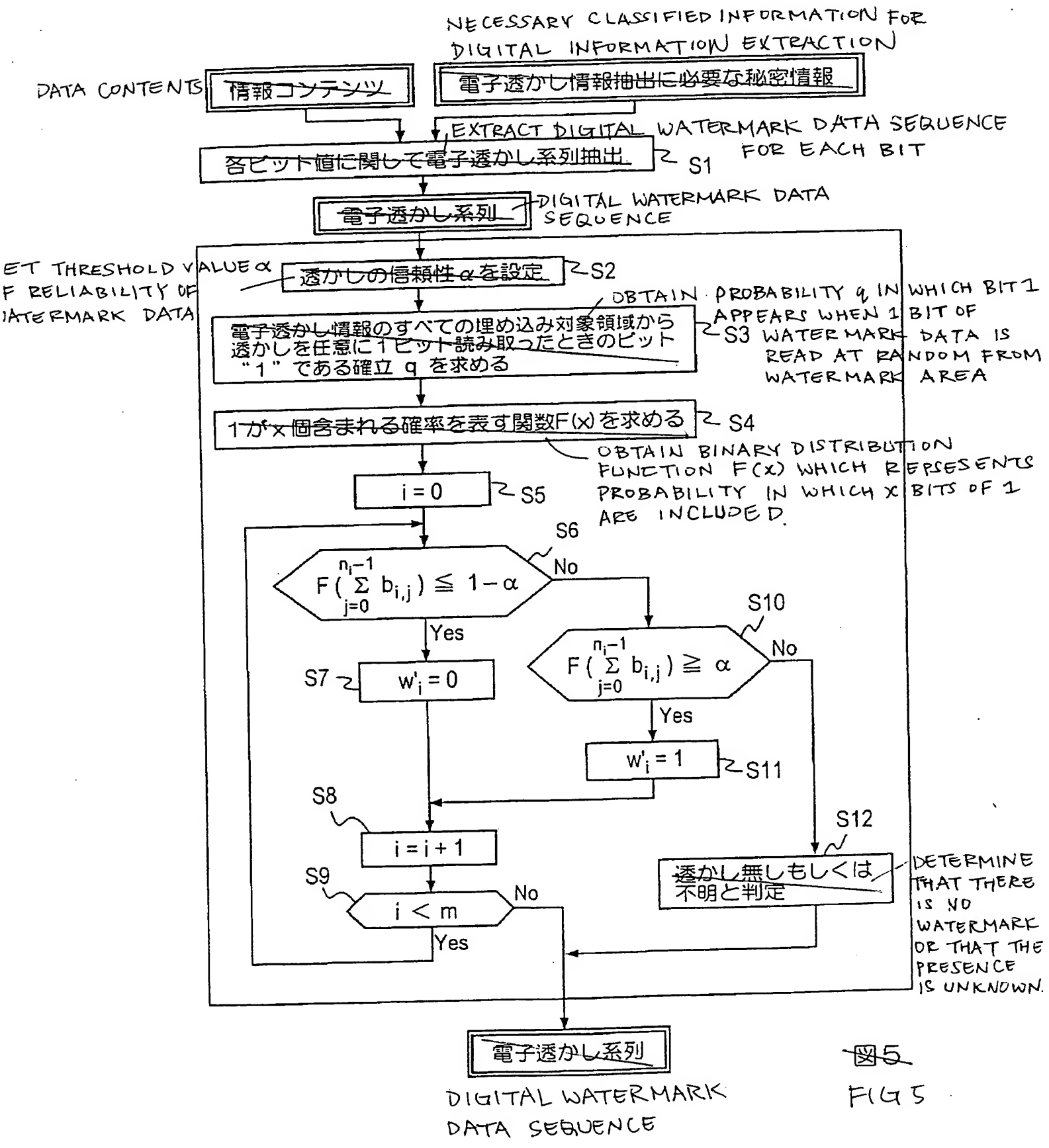


図5

FIG 5

【図６】
FIG. 6

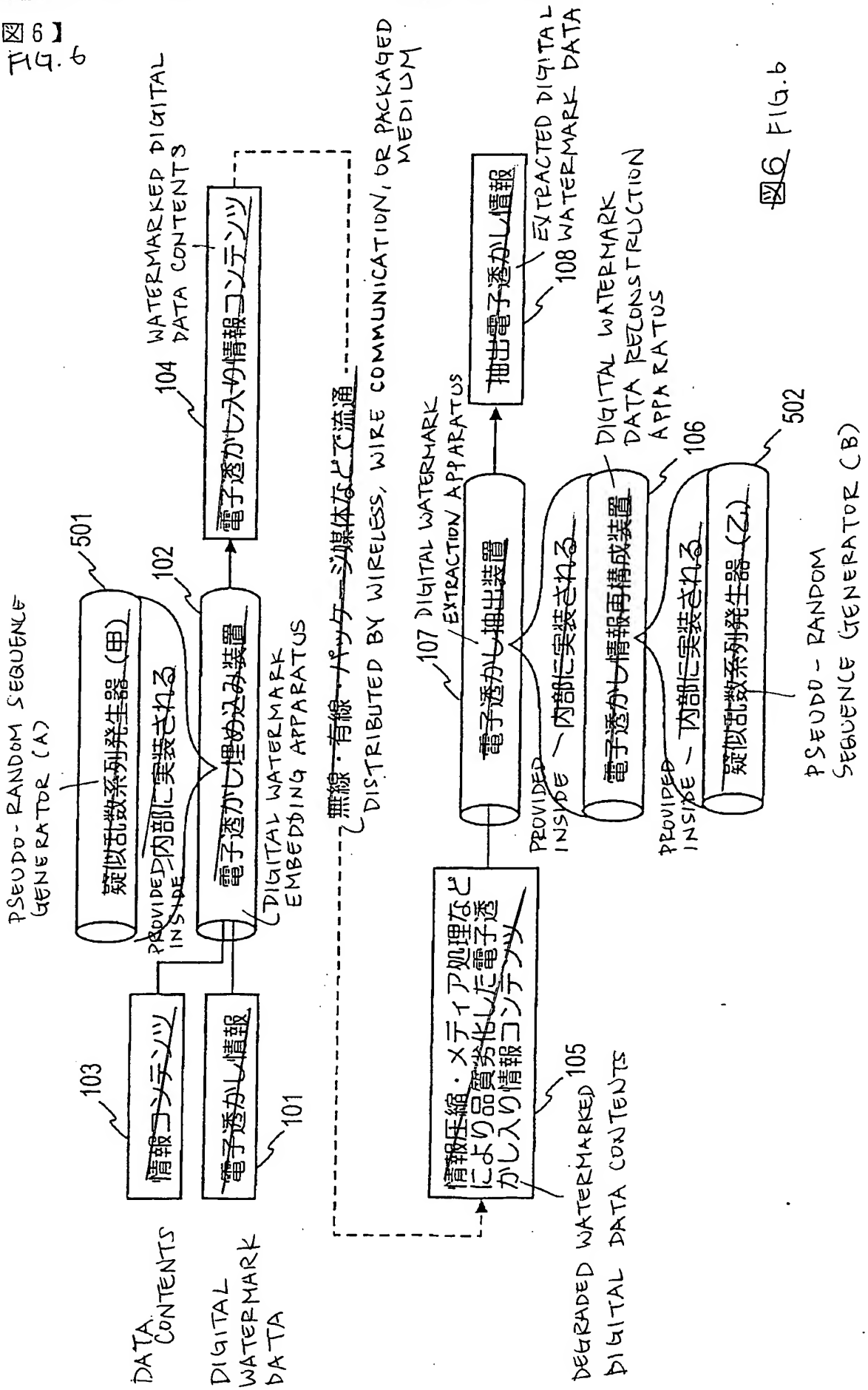


図6 FIG. 6

FIG.7 【図7】

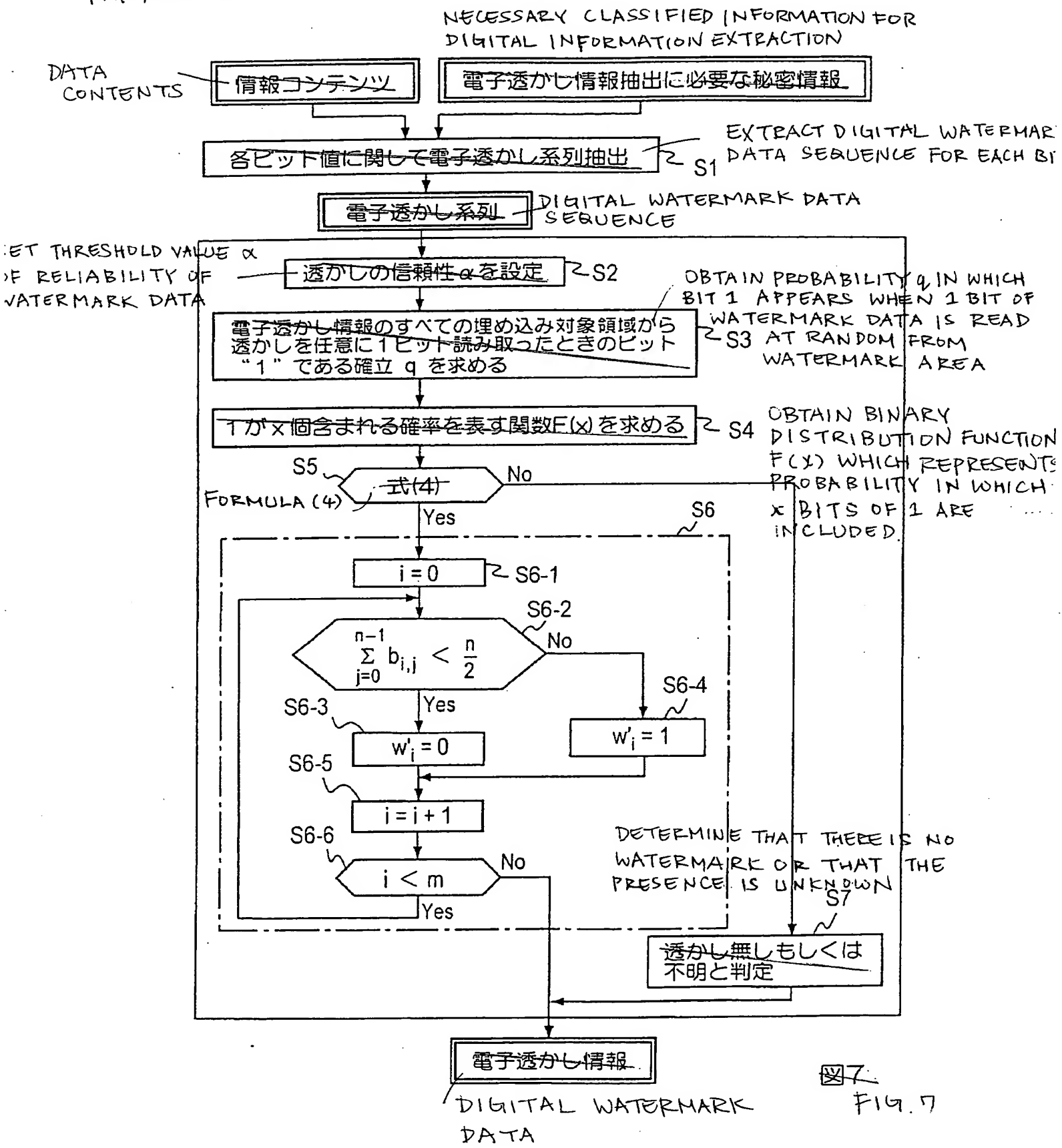


図7
FIG.7

【図8】
FIG. 8

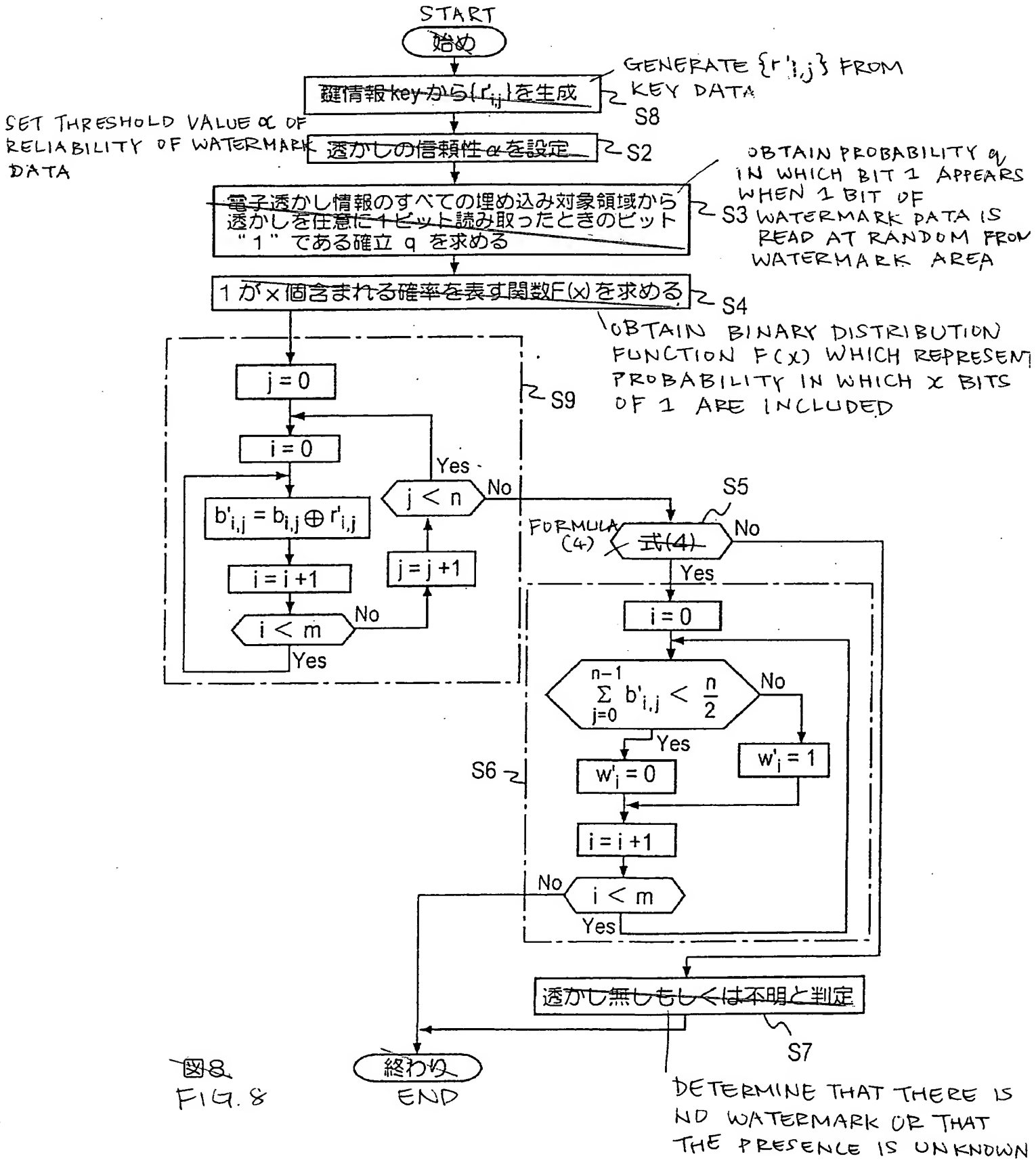


図8
FIG. 8

【図9】FIG.9

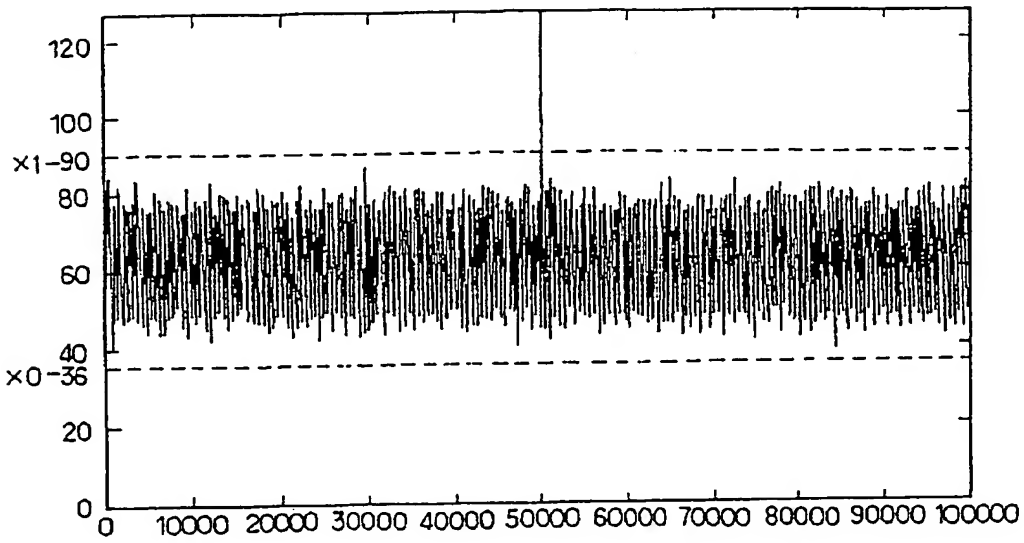


図9 FIG.9

【図10】FIG.10

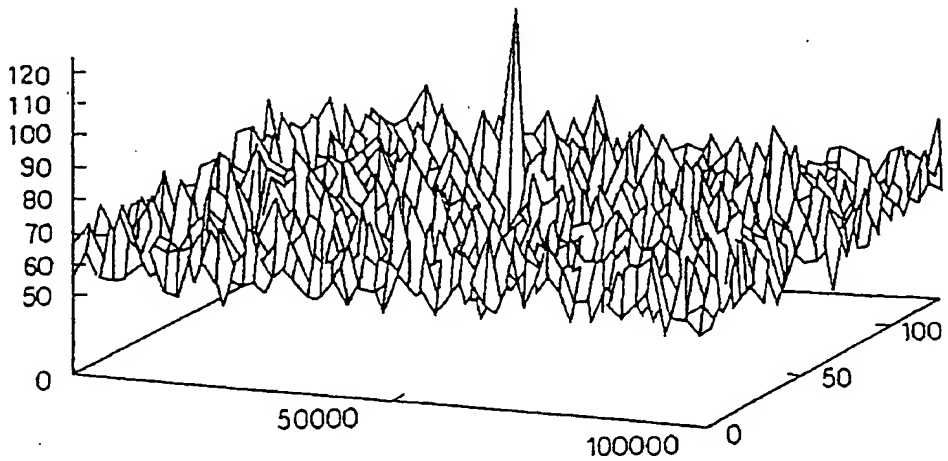


図10
FIG.10

BEST AVAILABLE COPY